A New Variable Regularized Transform Domain NLMS Adaptive Filtering Algorithm—Acoustic Applications and Performance Analysis

S. C. Chan, *Member, IEEE*, Y. J. Chu, and Z. G. Zhang, *Member, IEEE*

***Abstract*—This paper proposes a new regularized transform domain normalized LMS (R-TDNLMS) algorithm and studies its mean and mean square convergence performances. The proposed algorithm extends the conventional TDNLMS algorithm by imposing a regularization term on the filter coefficients to reduce the variance of estimators due to the lacking of excitation in a certain frequency band or in the presence of modeling errors. Difference equations describing the mean and mean square convergence behaviors of this algorithm are derived so as to characterize its convergence condition and steady-state excess mean square error (MSE). It shows that regularization can help to reduce the MSE by trading slight bias for variance. Based on this analysis, a new formula to select the regularization parameter for white Gaussian inputs is proposed, which leads to a new variable regularized TDNLMS (VR-TDNLMS) algorithm. Computer simulations are conducted to examine the improved convergence performance, steady-state MSE and robustness to power-varying inputs of the proposed algorithm and verify the effectiveness of the theoretical analysis. Furthermore, the application of the proposed VR-TDNLMS algorithm to the design and implementation of acoustic system identification and active noise control (ANC) systems show that they considerably outperforms traditional TDNLMS algorithms at low excitation or in the presence of modeling errors. Moreover, the theoretical analysis provides simple design formulas for achieving a given excess MSE (EMSE) and step-size bound for stable operation.**

***Index Terms*—Performance analysis, transform domain, variable regularization.**

I. Introduction

ADAPTIVE filters are frequently used in acoustic applications such as system identification, acoustic echo cancellation (AEC), active noise control (ANC) and related problems. Both the recursive least squares (RLS) algorithm [1] and the well known least mean square (LMS) algorithm [1], [2] as well as their variants [3]–[9] are commonly used. Due to numerical stability and computational simplicity, the LMS has received much attention in acoustic applications. An important class of the LMS algorithm is the transform domain normalized LMS (TDNLMS) algorithm [3]–[9], which exploits the decorrelation properties of transformations, such as discrete Fourier transform (DFT), discrete cosine transform (DCT), and wavelet transform (WT), to approximately whiten the input signal. This helps to reduce the eigenvalue spread of the input autocorrelation matrix and hence significantly improves the convergence speed of the conventional LMS algorithm.

An important problem of the TDNLMS and other LMS algorithms is their sensitivity to the level of the excitation signal, which may vary significantly over time as in speech and other audio signals. At low excitation, the estimated power of each transformed coefficient may become very small and the mean square errors (MSE) may increase significantly due to normalization of the transform coefficients by their estimated power. Another problem is its sensitivity to modelling errors, which is frequently encountered in systems such as ANC. In [7], a small constant is introduced to avoid numerical instabilities when the estimated input power is close to zero. Alternatively, a commonly used technique to address this issue is to introduce some kind of regularization into these algorithms. For example, if an L_2 regularization term on the filter coefficient is incorporated into the MSE cost function of conventional LMS algorithm, one obtains the classical Leaky LMS algorithm [1]. A constrained transform domain adaptive IIR filter structure for ANC was also proposed in [8], where soft and hard constraints analogous to regularization were applied to different transformed coefficients of a direct form adaptive IIR filter to mitigate the instability problem of the conventional filtered-U adaptive algorithm due to the poor frequency response of loudspeakers in ANC systems. Unlike the conventional TDNLMS algorithm, individual normalization of the transformed coefficient was not performed and hence it may still be sensitive to the eigenvalue spread problem of the conventional LMS algorithm. L_2 regularization has also been widely used in RLS algorithms [10], [11]. Moreover, the performances of the weighted L_2-based RLS and recursive least M-estimation algorithms have been analyzed in [11]. L_1 regularization, on the other hand, tends to produce sparse solutions as in [12], while the smoothly clipped absolute deviation (SCAD) [13] regularization and additional sparse enhancing transformations [14] have the advantage of asymptotic unbiased solution.

In this paper, a new regularized TDNLMS (R-TDNLMS) algorithm is proposed. A weighted L_2 regularization term on the adaptive filter coefficients is first incorporated in the MSE cost function in order to reduce the estimation variance. To quantify the performance of the proposed algorithm, its mean and

Manuscript received July 29, 2012; revised November 05, 2012 and November 20, 2012; accepted November 21, 2012. Date of publication November 30, 2012; date of current version January 23, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Lauri Savioja.

The authors are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong (e-mail: scchan@eee.hku.hk; yjchu@eee.hku.hk; zgzhang@eee.hku.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASL.2012.2231074

mean square convergence behaviors for Gaussian input and additive noise are analyzed. Difference equations describing the mean and mean square convergence behaviors of the proposed algorithm are derived. According to these difference equations, the convergence of these equations is analyzed by a new approach based on Lyapunov function. Analytical expressions for the steady-state excess MSE (EMSE) are derived. It is shown that the L_2 regularization introduces a bias to the Wiener solution. On the other hand, the regularization term helps to decrease the estimation variance at low input excitation and improve the convergence speed, especially for colored input. Moreover, by analyzing the effect of the L_2 regularization parameter on the steady-state MSE of the weight vector for white Gaussian inputs, a new formula for selecting the regularization parameter is obtained. This gives rise to the proposed variable regularized TDNLMS (VR-TDNLMS) algorithm.

Simulation results show that the proposed VR-TDNLMS algorithm has faster convergence rate and lower EMSE than the conventional TDNLMS algorithm. The theoretical analysis is also found to agree well with the simulation results. Furthermore, the application of the proposed algorithm to several acoustic applications including the design and implementation of acoustic system identification and ANC systems are studied. Simulation results show that the proposed VR-TDNLMS algorithm has better immunity to the variation in input signal power and robustness to modeling errors. Moreover, the theoretical analysis provides simple design formulas for achieving a given EMSE and step-size bound for stable operation.

The rest of the paper is organized as follows: in Section II, the R-TDNLMS algorithm is derived. The mean and mean square convergence behaviors of the proposed algorithm are then derived in Section III, from which the proposed regularization parameter selection method is obtained. The application of the proposed VR-TDNLMS algorithm to the system identification and ANC system is studied in Section IV. Finally, conclusions are drawn in Section V.

II. The Proposed R-TDNLMS Algorithm

Consider the identification of an L-order linear time-invariant (LTI) finite impulse response (FIR) system with coefficient vector $\boldsymbol{w}_0$ by an adaptive filter of the same length $\boldsymbol{w}(n) = [w_1(n), w_2(n), \ldots, w_L(n)]^T$. The unknown system and adaptive filter are both excited by an input $x(n)$ and the measured output of the system is $d(n)$, which is assumed to be corrupted by an additive noise $\eta(n)$, i.e.

$$d(n) = \boldsymbol{w}_0^T \boldsymbol{x}(n) + \eta(n), \tag{1}$$

where $\boldsymbol{x}(n) = [x(n), x(n-1), \ldots, x(n-L+1)]^T$ is the input signal vector. The adaptive filter aims to minimize a certain error measure of the estimation error $e(n) = d(n) - y(n)$, where $y(n) = \boldsymbol{w}^T(n)\boldsymbol{x}(n)$ is the output of the adaptive filter.

In conventional adaptive filtering, the mean square errors, $\varepsilon_{MSE} = E[e^2(n)]$, is minimized. Usually, ε_{MSE} is approximated by an exponentially weighted least square error function given by $J_{LS}(n) = \sum_{i=1}^{n} \lambda^{n-i} e^2(i)$, where λ is a positive forgetting factor between 0 and 1. In L_2 regularization [10], [11], a regularization term on the coefficient weight vector is introduced in order to reduce the variance of the estimator especially when the covariance matrix of $\boldsymbol{x}(n)$ is close to singular due to lacking of excitation. Then, the corresponding objective function becomes $J_{R\_LS}(n) = \sum_{i=1}^{n} \lambda^{n-i} e^2(i) + \xi \|\boldsymbol{w}(n)\|_2^2$, where ξ is a non-negative regularization parameter. Other popular regularization methods include the L_1 and SCAD regularization, which help to reduce the variance for system with sparse impulse response [13], [14].

In this paper, we shall consider the weighted L_2 regularization because it can be used to approximate the L_1 and SCAD regularization by changing the weight appropriately over time [14], [15].

Efficient batch algorithms for solving the L_2 and L_1 regularization problems have been addressed in a number of previous works such as [16]. In this work, we are interested in the recursive implementation similar to adaptive filtering for online applications. In particular, we shall focus on the development of a new regularized TDNLMS algorithm and the selection of the regularization parameter.

To start with, we consider the following weighted L_2 regularization problem

$$J_{WL\_2}(n) = \sum_{i=1}^{n} \lambda^{n-i} e^2(i) + \xi \|\boldsymbol{D}_W \boldsymbol{C}\boldsymbol{w}\|_2^2, \tag{2}$$

where $\boldsymbol{C}$ is an orthogonal transformation and $\boldsymbol{D}_W$ is a positive diagonal matrix. It will become apparent that the regularization penalizes the transformed weight vector $\boldsymbol{C}\boldsymbol{w}$ with large amplitude. $\boldsymbol{D}_W$ can be made adaptive so as to approximate different regularization methods. For instance, $\boldsymbol{D}_W(n)$ is an identity matrix for L_2 regularization or the generalized inverse of the diagonal matrix $diag\{\sqrt{|W_1(n-1)|}, \cdots, \sqrt{|W_L(n-1)|}\}$, where $\boldsymbol{W}(n) = \boldsymbol{C}\boldsymbol{w}(n) = [W_1(n), \ldots, W_L(n)]^T$ for L_1 regularization. Note, $\boldsymbol{C}$ serves as a sparsity-enhancing transformation, which may improve further the performance of the system [14].

A first order necessary condition for optimality is $\nabla_w(J_{WL\_2}(n)) = \boldsymbol{0}$, where $\boldsymbol{0}$ is a null vector. Since the problem is convex, this is also the global optimal condition. Since the gradient of (2) with respect to $\boldsymbol{w}$ is

$$\nabla_w \left(J_{WL\_2}(n)\right) = -2 \sum_{i=1}^{n} \lambda^{n-i} e(i)\boldsymbol{x}(i) + 2\xi \boldsymbol{C}^T \boldsymbol{R}_W \boldsymbol{C}\boldsymbol{w}, \tag{3}$$

where $\boldsymbol{R}_W = \boldsymbol{D}_W^T \boldsymbol{D}_W$ is a regularization matrix, and the optimal solution $\boldsymbol{w}_R$ satisfies the following system of linear equation:

$$\boldsymbol{G}_{xx}(n)\boldsymbol{w}_R = \boldsymbol{r}_{xd}(n), \tag{4}$$

where $\boldsymbol{r}_{xd}(n) = \sum_{i=1}^{n} \lambda^{n-i} d(i)x(i)$ is the cross-correlation vector, and $\boldsymbol{G}_{xx}(n) = \boldsymbol{R}_{xx}(n) + \xi \boldsymbol{C}^T \boldsymbol{R}_W \boldsymbol{C}$ is the regularized covariance matrix of $x(n)$ with $\boldsymbol{R}_{xx}(n) = \sum_{i=1}^{n} \lambda^{n-i} \boldsymbol{x}(i)\boldsymbol{x}(i)^T$ the covariance matrix of $x(n)$. The problem is also referred to as ridge regression in statistics and it can be computed by solving the system of linear equation above.

The above problem can also be solved by the Newton's method given any vector $\boldsymbol{w}(n)$. The next iteration is given by

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) - \frac{1}{2}\mu \boldsymbol{G}_{xx}^{-1}(n) \nabla_w \left(J_{WL\_2}(n)\right)|_{\boldsymbol{w}(n)}, \tag{5}$$

where μ is the step-size parameter and $\nabla_w(J_{WL\_2}(n))|_{\boldsymbol{w}(n)}$ is the gradient evaluated at $\boldsymbol{w}(n)$. For the weighted L_2 problem, the Newton method will converge in one step with $\mu = 1$. The remaining problem is then to compute $\boldsymbol{G}_{xx}^{-1}(n)$ recursively over time. Although the change in $\boldsymbol{R}_{xx}(n)$ over time is rank-one, the possible change in $\xi \boldsymbol{C}^T \boldsymbol{R}_W \boldsymbol{C}$ is full rank. One simplification is to decompose the full rank regularization $\boldsymbol{D}_W \boldsymbol{C}$ into rank-one rows and apply them successively to the QR decomposition to realize full-rank regularization with a given strength [10], [14]. This gives rise to a numerically better behaved QR decomposition-based algorithm.

In this paper, we seek to simplify the implementation further with the aim to derive a TDNLMS-like algorithm. The key is to assume that an appropriate orthogonal transformation can be chosen such that $\boldsymbol{C}\boldsymbol{G}_{xx}^{-1}(n)\boldsymbol{C}^T \approx \boldsymbol{D}_x$, where $\boldsymbol{D}_x$ is a diagonal matrix. In other words, $\boldsymbol{G}_{xx}^{-1}(n)$ is approximately diagonalized by $\boldsymbol{C}$ and consequently we only need to estimate $\boldsymbol{D}_x$. The matrix $\boldsymbol{C}$ can be chosen as the DCT, DFT etc because of their good decorrelation property. Using this property and pre-multiplying (5) by $\boldsymbol{C}$, (5) can be simplified to

$$\begin{aligned}\boldsymbol{W}(n+1) &= \boldsymbol{W}(n) - \frac{1}{2}\mu\left(\boldsymbol{C}\boldsymbol{G}_{xx}^{-1}(n)\boldsymbol{C}^T\right)\left(\boldsymbol{C}\nabla_w\left(J_{WL\_2}(n)\right)\right)\\ &\approx \boldsymbol{W}(n) - \frac{1}{2}\mu\boldsymbol{D}_x\left(\boldsymbol{C}\nabla_w\left(J_{WL\_2}(n)\right)\right),\end{aligned}$$

where $\boldsymbol{W}(n) = \boldsymbol{C}\boldsymbol{w}(n)$. Moreover, by employing the fact

$$\boldsymbol{C}\nabla_w\left(J_{WL\_2}(n)\right) = -2\sum_{i=1}^{n}\lambda^{n-i}e(i)\boldsymbol{X}_C(n) + 2\xi\boldsymbol{R}_W\boldsymbol{W}(n),$$

one obtains the following simplified update equation

$$\boldsymbol{W}(n+1) = (\boldsymbol{I} - \mu\xi\boldsymbol{D}_x\boldsymbol{R}_W)\boldsymbol{W}(n) + \mu\boldsymbol{D}_x\sum_{i=1}^{n}\lambda^{n-i}e(i)\boldsymbol{X}_C(n), \tag{6}$$

where $\boldsymbol{X}_C(n) = \boldsymbol{C}\boldsymbol{x}(n)$ is the transformed input vector, and $\boldsymbol{I}$ is the identity matrix. Furthermore, if one approximates the weighted error $\sum_{i=1}^{n}\lambda^{n-i}e(i)\boldsymbol{X}_C(n)$ using the instantaneous error, (6) will reduce to

$$\boldsymbol{W}(n+1) = (\boldsymbol{I} - \mu\xi\boldsymbol{D}_x\boldsymbol{R}_W)\boldsymbol{W}(n) + \mu\boldsymbol{D}_x\boldsymbol{X}_C(n)e(n). \tag{7}$$

Next, we examine the approximation of the diagonal matrix $\boldsymbol{D}_x \approx \boldsymbol{C}\boldsymbol{G}_{xx}^{-1}(n)\boldsymbol{C}^T$ or in other words $\boldsymbol{G}_{xx}(n) \approx \boldsymbol{C}^T\boldsymbol{D}_x^{-1}\boldsymbol{C}$. First, we note from (4) that $\boldsymbol{G}_{xx}(n) = \boldsymbol{R}_{xx}(n) + \xi\boldsymbol{C}^T\boldsymbol{R}_W\boldsymbol{C}$ and hence $\boldsymbol{G}_{xx}(n) = \boldsymbol{C}^T(\boldsymbol{C}\boldsymbol{R}_{xx}(n)\boldsymbol{C}^T + \xi\boldsymbol{R}_W)\boldsymbol{C} \approx \boldsymbol{C}^T\boldsymbol{D}_x^{-1}\boldsymbol{C}$. If the transform matrix $\boldsymbol{C}$ approximately diagonalizes $\boldsymbol{R}_{xx}(n)$, then $\boldsymbol{C}^T(\Lambda_x + \xi\boldsymbol{R}_W)\boldsymbol{C} \approx \boldsymbol{C}^T\boldsymbol{D}_x^{-1}\boldsymbol{C}$ and $\varepsilon_i^{-1} \approx [\Lambda_x]_i + \xi[\boldsymbol{R}_W]_{i,i}$, where ε_i^{-1} is the i-th diagonal element of $\boldsymbol{D}_x^{-1}$, $[\Lambda_x]_i$ is the i-th eigenvalue of $\boldsymbol{R}_{xx}(n)$, and $[\boldsymbol{R}_W]_{i,i}$ denotes the (i,i) element of $\boldsymbol{R}_W$. Since $[\boldsymbol{C}\boldsymbol{R}_{xx}\boldsymbol{C}^T]_{i,i} \approx E[X_{C\_i}^2(n)]$, which can be estimated recursively as

$$\sigma_{X_{C\_i}}^2(n+1) = \lambda\sigma_{X_{C\_i}}^2(n) + (1-\lambda)X_{C\_i}^2(n),$$

and $\boldsymbol{R}_{xx}(n)$ for large n is equal to $(1/(1-\lambda))E[\boldsymbol{x}(n)\boldsymbol{x}^T(n)] = (1/((1-\lambda))\boldsymbol{R}_{xx}$, we have $[\Lambda_x]_i \approx (1/((1-\lambda))\sigma_{X_{C\_i}}^2(n)$. For convenience, one can absorb the scaling $1/(1-\lambda)$ into μ and ξ. Thus the estimate $[\boldsymbol{D}_x]_i$ in (7) can be simplified as

$$\varepsilon_i(n) \approx \left(\sigma_{X_{C\_i}}^2(n) + \xi[\boldsymbol{R}_W]_{i,i}\right)^{-1}, \tag{8}$$

Consequently, (7) and (8) constitute the proposed R-TDNLMS algorithm. For simplicity, we shall consider the case with a fixed step-size μ. Other possible and simpler update of $\varepsilon_i(n)$ is

$$\varepsilon_i(n) \approx \left(\varepsilon + \sigma_{X_{C\_i}}^2(n)\right)^{-1}, \tag{9}$$

where ε is a positive constant. If $\boldsymbol{C} = \boldsymbol{I}$, one gets the regularized LMS algorithm as follows:

$$\boldsymbol{w}(n+1) = (\boldsymbol{I} - \mu\xi\boldsymbol{R}_W)\boldsymbol{w}(n) + \mu\boldsymbol{x}(n)e(n). \tag{10}$$

When $\boldsymbol{D}_x = \boldsymbol{R}_W = \boldsymbol{I}$, (10) reduces to the familiar Leaky LMS algorithm [1]. The matrix $\boldsymbol{D}_W$ can be chosen from the current weight vector to approximate L_1 or SCAD regularization, which generalizes the classical Leaky LMS algorithm. However, it is known that the convergence speed of the conventional LMS algorithm is sensitive to eigenvalue spread of the input covariance matrix $E[\boldsymbol{x}(n)\boldsymbol{x}^T(n)]$. The proposed regularized TDNLMS approximately whitens the input by means of an appropriate orthogonal transformation and element-wise normalization of the transformed coefficients, which leads to an improved performance. The proposed R-TDNLMS differs from the constrained transform domain adaptive filter in [8] in that the latter focus on direct form IIR filter with $\boldsymbol{D}_x = \boldsymbol{R}_W = \boldsymbol{I}$, which makes its basic form identical to the Leaky LMS algorithm, except that the transformation $\boldsymbol{C}$ is a skinny $M \times L$ matrix with $L > M$, so that some of the transformed coefficients which lie at the poor frequency bands of the speaker are simply ignored to avoid instability. This removal can be viewed as hard constraints, while the regularization serves as soft constraints of the system to improve system stability. However, since the transformed coefficients are not normalized, it will still be sensitive to possible eigenvalue spread of the input covariance matrix. Moreover, it cannot be used directly to implement approximate L_1 and SCAD regularizations.

Unlike most previous works which employ a constant regularization parameter, we propose to make ξ adaptive, which gives rise to a new variable regularized TDNLMS (VR-TDNLMS) algorithm. In order to derive such a parameter selection method to be discussed in Section III-C, we first examine the convergence behaviors and steady-state EMSE of the R-TDNLMS algorithm.

III. Performance Analysis

In this section, the convergence performance of the proposed R-TDNLMS algorithm with Gaussian input and additive noises is analyzed. The following assumptions are made:

(A1) $\{\boldsymbol{x}(n)\}$ is Gaussian distributed sequence with zero mean and covariance matrix $\boldsymbol{R}_{xx}$;

(A2) $\boldsymbol{W}(n)$, $\boldsymbol{x}(n)$ and $\eta(n)$ are statistically independent;

(A3) $\eta(n)$ is white Gaussian distributed with zero mean and variance σ_η^2.

(A4) the elements in $\boldsymbol{D}_x$, $\varepsilon_i(n)$, are uncorrelated with $\boldsymbol{W}(n)$ and $\boldsymbol{x}(n)$ due to the recursive averaging effect of $\sigma_{X_{C\_i}}^2$ in (8).

(A2) is the commonly used independence assumption, which is a good approximation for large L and for small to medium step-sizes to simplify the convergence analysis of adaptive algorithms. Moreover, we denote the optimal transformed weight vector by $\boldsymbol{W}_0 = \boldsymbol{R}_{X_C X_C}^{-1} \boldsymbol{P}_{dX_C}$, where $\boldsymbol{R}_{XcXc} = E[\boldsymbol{X}_C(n)\boldsymbol{X}_C^T(n)]$ is the transform domain input covariance matrix and $\boldsymbol{P}_{dXc} = E[d(n)\boldsymbol{X}_C(n)]$ is the cross-correlation vector between $\boldsymbol{X}_C(n)$ and $d(n)$. $\boldsymbol{W}_0$ is related to the optimal Wiener solution $\boldsymbol{w}_0$ by $\boldsymbol{w}_0 = \boldsymbol{R}_{X_C X_C}^{-1} \boldsymbol{P}_{dX} = \boldsymbol{C}\boldsymbol{W}_0$.

A. Mean Convergence Analysis

We now present the mean convergence behavior of the proposed algorithm. First, we assume that the algorithm is convergent. The condition of convergence will be shown later in this section. Taking expectation on both sides of (7), we have

$$E[\boldsymbol{W}(n+1)] = E[\boldsymbol{W}(n)] + \mu \boldsymbol{D}_x \cdot \{\boldsymbol{P}_{dX_C} - (\boldsymbol{R}_{X_C X_C} + \xi \boldsymbol{R}_W) E[\boldsymbol{W}(n)]\}. \quad (11)$$

At the steady state, (11) reads

$$\boldsymbol{W}_R = (\boldsymbol{R}_{X_C X_C} + \xi \boldsymbol{R}_W)^{-1} \boldsymbol{P}_{dX_C}, \quad (12)$$

where $\boldsymbol{W}_R = E[\boldsymbol{W}(\infty)]$ is the desired transformed regularized Wiener solution. The bias introduced can be estimated by rewriting (12) as

$$\left(\boldsymbol{I} + \xi \boldsymbol{R}_{X_C X_C}^{-1} \boldsymbol{R}_W\right) \boldsymbol{W}_R = \boldsymbol{R}_{X_C X_C}^{-1} \boldsymbol{P}_{dX_C} = \boldsymbol{W}_0. \quad (13)$$

Let $\Delta \boldsymbol{W} = \boldsymbol{W}_0 - \boldsymbol{W}_R$ be the bias from the Wiener solution. Using the expansion $(\boldsymbol{I} - \boldsymbol{P})^{-1} = \sum_{k=0}^{\infty} \boldsymbol{P}^k$ for $\boldsymbol{P}$ with spectral radius less than 1, one gets

$$\Delta \boldsymbol{W} = \boldsymbol{W}_0 - \boldsymbol{W}_R = \boldsymbol{W}_0 - \sum_{k=0}^{\infty} \left(-\xi \boldsymbol{R}_{X_C X_C}^{-1} \boldsymbol{R}_W\right)^k \boldsymbol{W}_0. \quad (14)$$

Here, we have assumed that the spectral radius of $\xi \boldsymbol{R}_{X_C X_C}^{-1} \boldsymbol{R}_W$ is less than 1. This is valid if ξ is sufficiently small and $\boldsymbol{R}_{X_C X_C}$ is nonsingular. For mild regularization, ξ is small and a first order approximation of the bias is $\Delta \boldsymbol{W} \approx \xi \boldsymbol{R}_{X_C X_C}^{-1} \boldsymbol{R}_W \boldsymbol{W}_0$.

Next we examine the convergence rate by introducing the weight error vector $\boldsymbol{v}(n) = \boldsymbol{W}(n) - \boldsymbol{W}_R(n)$ in (11), which yields

$$E[\boldsymbol{v}(n+1)] = [\boldsymbol{I} - \mu \boldsymbol{D}_x (\boldsymbol{R}_{X_C X_C} + \xi \boldsymbol{R}_W)] E[\boldsymbol{v}(n)]. \quad (15)$$

Let $\widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{U}}^T$ be the eigendecomposition of $\overline{\boldsymbol{R}}_{xx} = \boldsymbol{D}_x^{1/2} \widetilde{\boldsymbol{R}}_{X_C X_C} \boldsymbol{D}_x^{1/2}$ with $\widetilde{\boldsymbol{R}}_{X_C X_C} = \boldsymbol{R}_{X_C X_C} + \xi \boldsymbol{R}_W$. Using (15) and the change of variable $\boldsymbol{V}(n) = \widetilde{\boldsymbol{U}}^T \boldsymbol{D}_x^{-1/2} \boldsymbol{v}(n)$, we get the difference equation for the i-th element of $E[\boldsymbol{V}(n)]$ as follows

$$E[\boldsymbol{V}(n+1)]_i = (1 - \mu \widetilde{\lambda}_i) E[\boldsymbol{V}(n)]_i, \quad (16)$$

where $\widetilde{\lambda}_i$ is the i-th eigenvalue of $\overline{R}_{xx}$. Thus, the mean weight vector of the adaptive filter will converge if

$$0 < \mu < \frac{2}{\widetilde{\lambda}_i}. \quad (17)$$

Therefore, the maximum possible step-size is $\mu_{\max} = 2/\widetilde{\lambda}_{\max}$, where $\widetilde{\lambda}_{\max}$ is the maximum eigenvalue of $\overline{\boldsymbol{R}}_{\boldsymbol{xx}}$. Moreover, since the maximum eigenvalue of $\widetilde{\boldsymbol{R}}_{X_C X_C} = \boldsymbol{R}_{X_C X_C} + \xi \boldsymbol{R}_W$ is larger than that of $\boldsymbol{R}_{X_C X_C}$ for positive ξ, $\widetilde{\lambda}_{\max}$ will be larger than the maximum eigenvalue of the conventional covariance matrix $\boldsymbol{R}_{X_C X_C}$ in the TDNLMS algorithm. Therefore, the use of regularization will reduce slightly the maximum possible step-size.

It can be also seen that if the input covariance matrix $\boldsymbol{R}_{X_C X_C}$ has zero eigenvalues, these eigenmodes can never converge and the solution may be significantly biased. With sufficient regularization, the eigenvalues of $\overline{\boldsymbol{R}}_{xx}$ can be made nonzero and hence the solution will be given by (12) with a controllable bias.

B. Mean Square Convergence Analysis

To evaluate the mean square behavior, we multiply $\boldsymbol{v}(n)$ in (7) by its transpose and take expectation to obtain a difference equation for the weight error covariance matrix $\boldsymbol{\Xi}_{vv}(n) = E[\boldsymbol{v}(n)\boldsymbol{v}^T(n)]$ as follows

$$\begin{aligned}\boldsymbol{\Xi}_{vv}(n+1) &= \boldsymbol{\Xi}_{vv}(n) - \mu\left(\boldsymbol{D}_x \tilde{\boldsymbol{R}}_{X_C X_C} \boldsymbol{\Xi}_{vv}(n) + \boldsymbol{\Xi}_{vv}(n) \tilde{\boldsymbol{R}}_{X_C X_C} \boldsymbol{D}_x^T\right) \\ &\quad + \mu^2 \boldsymbol{D}_x E\Big[\left(\boldsymbol{X}_C(n)e(n) - \xi \boldsymbol{R}_W \boldsymbol{W}(n)\right) \\ &\quad \cdot \left(\boldsymbol{X}_C^T(n)e(n) - \xi \boldsymbol{W}^T(n)\boldsymbol{R}_W\right)\Big] \boldsymbol{D}_x^T. \quad (18)\end{aligned}$$

By noting that $e(n) = [\boldsymbol{W}_0 - \boldsymbol{W}(n)]^T \boldsymbol{X}_C(n) + \eta(n) = \hat{\boldsymbol{v}}^T \boldsymbol{X}_C(n) + \eta(n)$, where $\hat{\boldsymbol{v}} = \boldsymbol{W}_0 - \boldsymbol{W}_R - \boldsymbol{v}(n) = \Delta \boldsymbol{W} - \boldsymbol{v}(n)$, the last term in (18) can be rewritten as

$$\boldsymbol{S}(n) = \mu^2 \boldsymbol{D}_x [\boldsymbol{A}_0(n) - \boldsymbol{A}_1(n) - \boldsymbol{A}_2(n) + \boldsymbol{A}_3(n) + \boldsymbol{R}_{X_C X_C} \sigma_\eta^2] \boldsymbol{D}_x^T,$$

where the expectations are summarized for further analysis as follows $\boldsymbol{A}_0(n) = E[\boldsymbol{X}_C(n)\boldsymbol{X}_C^T(n)\boldsymbol{\Xi}_{\hat{v}\hat{v}}(n)\boldsymbol{X}_C(n)\boldsymbol{X}_C^T(n)]$, $\boldsymbol{A}_1(n) = \boldsymbol{A}_2^T(n) = \xi \boldsymbol{R}_{X_C X_C} \boldsymbol{\Xi}_{\hat{v}w}(n) \boldsymbol{R}_W$ and $\boldsymbol{A}_3(n) = \xi^2 \boldsymbol{R}_{\boldsymbol{W}} \boldsymbol{\Xi}_{ww}(n) \boldsymbol{R}_{\boldsymbol{W}}$ with $\boldsymbol{\Xi}_{\hat{v}\hat{v}}(n) = \Delta \boldsymbol{W} \Delta \boldsymbol{W}^T + \boldsymbol{\Xi}_{vv}(n) - \overline{\boldsymbol{v}}(n) \Delta \boldsymbol{W}^T - \Delta \boldsymbol{W} \overline{\boldsymbol{v}}^T(n)$, $\boldsymbol{\Xi}_{ww}(n) = \boldsymbol{W}_R \boldsymbol{W}_R^T + \boldsymbol{\Xi}_{vv}(n) + \overline{\boldsymbol{v}}(n) \boldsymbol{W}_R^T + \boldsymbol{W}_R \overline{\boldsymbol{v}}^T(n)$, $\boldsymbol{\Xi}_{\hat{v}w}(n) = \Delta \boldsymbol{W} \boldsymbol{W}_R^T - \boldsymbol{\Xi}_{vv}(n) + \overline{\boldsymbol{v}}(n) \boldsymbol{W}_R^T - \Delta \boldsymbol{W} \overline{\boldsymbol{v}}^T(n)$, and $\overline{\boldsymbol{v}}(n) = E[\boldsymbol{v}(n)]$.

Next, we consider the stability of the algorithm.

Stability and Step-Size Bound: Multiplying both sides of (18) by $\boldsymbol{D}_x^{-1}$ and taking the trace operation $\varphi(n) = Tr(\boldsymbol{D}_x^{-1} \boldsymbol{\Xi}_{vv}(n))$, we get

$$\varphi(n+1) = \varphi(n) - \Big[2\mu Tr\Big(\tilde{\boldsymbol{R}}_{X_C X_C} \boldsymbol{\Xi}_{vv}(n)\Big) - Tr\Big(\boldsymbol{D}_x^{-1} \boldsymbol{S}(n)\Big)\Big], \quad (19)$$

where $\varphi(n)$ serves the role of the Lyapunov function, which is always positive. Since the driving terms in (19) with respect to σ_η^2, $\boldsymbol{W}_R$, and $\Delta \boldsymbol{W}$ are finite, and $E[\boldsymbol{v}]$ converges to zero if (17) is satisfied, we only need to consider those terms in $\boldsymbol{\Xi}_{vv}(n)$. Moreover, as $\varphi(n)$ is positive, if the term inside the bracket in

(19) is positive, then the system is guaranteed to be stable. Inserting $\boldsymbol{A}_0(n)$, $\boldsymbol{A}_1(n)$, $\boldsymbol{A}_2(n)$ and $\boldsymbol{A}_3(n)$ into (18) and simplifying, one gets the following sufficient condition for stability:

$$2\mu Tr\left(\tilde{\boldsymbol{R}}_{X_C X_C}\boldsymbol{\Xi}_{vv}(n)\right) > \mu^2\gamma(n), \tag{20}$$

where we define the following term for simplicity

$$\begin{aligned}\gamma(n) = &\, 2Tr\left(\boldsymbol{R}_{X_C X_C}\boldsymbol{\Xi}_{vv}(n)\boldsymbol{R}_{X_C X_C}\boldsymbol{D}_x\right)\\ &+ Tr\left(\boldsymbol{R}_{X_C X_C}\boldsymbol{\Xi}_{vv}(n)\right)Tr\left(\boldsymbol{R}_{X_C X_C}\boldsymbol{D}_x\right)\\ &+ \xi Tr\left(\boldsymbol{R}_{X_C X_C}\boldsymbol{\Xi}_{vv}(n)\boldsymbol{R}_W\boldsymbol{D}_x\right)\\ &+ \xi Tr\left(\boldsymbol{R}_W\boldsymbol{\Xi}_{vv}(n)\boldsymbol{R}_{X_C X_C}\boldsymbol{D}_x\right)\\ &+ \xi^2 Tr\left(\boldsymbol{R}_W\boldsymbol{\Xi}_{vv}(n)\boldsymbol{R}_W\boldsymbol{D}_x\right).\end{aligned}$$

Here, we have used the Gaussian factor theorem to obtain:

$$\begin{aligned}&E\left[\boldsymbol{X}_C(n)\boldsymbol{X}_C^T(n)\boldsymbol{\Xi}_{vv}(n)\boldsymbol{X}_C(n)\boldsymbol{X}_C^T(n)\right]\\ &= 2\boldsymbol{R}_{X_C X_C}\boldsymbol{\Xi}_{vv}(n)\boldsymbol{R}_{X_C X_C} + Tr\left(\boldsymbol{R}_{X_C X_C}\boldsymbol{\Xi}_{vv}(n)\right)\boldsymbol{R}_{X_C X_C}\end{aligned}$$

so as to arrive at the first two terms of $\gamma(n)$ in (20).

Since, $\gamma(n)$ depends on $\boldsymbol{\Xi}_{vv}(n)$, (20) cannot be used directly to determine a bound on μ for practical application. Fortunately, we can use the facts that $Tr(\boldsymbol{AB}) \le Tr(\boldsymbol{A})Tr(\boldsymbol{B})$ for $\boldsymbol{A}, \boldsymbol{B} > 0$ and $Tr(\boldsymbol{AB}) = Tr(\boldsymbol{BA})$ to obtain an upper bound for $\gamma(n)$ as follows:

$$\gamma(n) \le Tr\left(\tilde{\boldsymbol{R}}_{X_C X_C}\boldsymbol{\Xi}_{vv}(n)\right)\cdot\widetilde{\gamma} \tag{21}$$

where

$$\begin{aligned}\widetilde{\gamma} = &\, 2Tr\left(\boldsymbol{R}_{X_C X_C}\boldsymbol{D}_x\Gamma_x\right) + Tr\left(\boldsymbol{R}_{X_C X_C}\boldsymbol{D}_x\right)Tr(\Gamma_x)\\ &+2\xi Tr(\boldsymbol{R}_W\boldsymbol{D}_x\Gamma_x) + \xi^2 Tr\left(\boldsymbol{R}_W\boldsymbol{D}_x\boldsymbol{R}_W\widetilde{\boldsymbol{R}}_{X_C X_C}^{-1}\right)\end{aligned}$$

and $\Gamma_x = \boldsymbol{R}_{X_C X_C}\widetilde{\boldsymbol{R}}_{X_C X_C}^{-1}$. By cancelling the term $Tr(\widetilde{\boldsymbol{R}}_{X_C X_C}\boldsymbol{\Xi}_{vv}(n))$ on both sides of (20) using (21), one gets the following conservative upper bound of μ for mean squares convergence:

$$\mu_{R-TDLMS} \le \frac{2}{\widetilde{\gamma}}. \tag{22}$$

For conventional TDLMS algorithm without regularization, one has $\Gamma_x = \boldsymbol{I}$ and $\xi = 0$, which yields the familiar step-size bound for mean square convergence:

$$\mu_{TDLMS} \le \frac{2}{3Tr\left(\boldsymbol{R}_{X_C X_C}\boldsymbol{D}_x\right)}.$$

Again, one notices that the regularization slightly reduces the step-size bound for mean square convergence as compared with the conventional algorithm. As we shall show later that the price paid by this reduction is rewarded by the reduced MSE if $\boldsymbol{R}_{X_C X_C}$ is ill-conditioned.

When $\boldsymbol{C} = \boldsymbol{D}_x = \boldsymbol{I}$, it will further reduce to the conventional LMS algorithm. It can be seen that (22) reduces to the classical results of Weinstein [17], which was obtained by solving the difference equations followed by certain simplification. It therefore suggests that the proposed stability analysis method based on the Lyapunov function and the trace operator is a very convenient and powerful tool for analyzing the stability and deriving the maximum possible step-size for adaptive filters. We now derive an expression for the steady-state EMSE.

Steady-State EMSE: If the algorithm converges, $E[\boldsymbol{v}(\infty)] = \boldsymbol{0}$, and the last term in (18) will reduce to

$$\begin{aligned}S(\infty) =&\, \mu^2\boldsymbol{D}_x\left[\boldsymbol{A}_0(\infty) - \boldsymbol{A}_1(\infty) - \boldsymbol{A}_2(\infty) + \boldsymbol{A}_3(\infty)\right.\\ &\left.+ \boldsymbol{R}_{X_C X_C}\sigma_\eta^2\right]\boldsymbol{D}_x^T,\\ \boldsymbol{A}_0(\infty) =&\, 2\boldsymbol{R}_{X_C X_C}\boldsymbol{\Xi}_{vv}(\infty)\boldsymbol{R}_{X_C X_C} + 2\xi^2\tilde{\boldsymbol{R}}_W\\ &+ Tr\left(\left[\Delta\boldsymbol{W}\Delta\boldsymbol{W}^T + \boldsymbol{\Xi}_{vv}(\infty)\right]\boldsymbol{R}_{X_C X_C}\right)\boldsymbol{R}_{X_C X_C},\\ \boldsymbol{A}_1(\infty) =&\, \boldsymbol{A}_2^T(\infty) = \xi\left(\xi\tilde{\boldsymbol{R}}_W - \boldsymbol{R}_{X_C X_C}\boldsymbol{\Xi}_{vv}(\infty)\boldsymbol{R}_W\right),\\ \boldsymbol{A}_3(\infty) =&\, \xi^2\left(\tilde{\boldsymbol{R}}_W + \boldsymbol{R}_W^T\boldsymbol{\Xi}_{vv}(\infty)\boldsymbol{R}_W\right),\\ &\text{where } \tilde{\boldsymbol{R}}_W = \boldsymbol{R}_{\boldsymbol{W}}\boldsymbol{W}_R\boldsymbol{W}_R^T\boldsymbol{R}_{\boldsymbol{W}}^T\end{aligned}$$

and we have used the identity $\boldsymbol{R}_{X_C X_C}\Delta\boldsymbol{W} = \xi\boldsymbol{R}_W\boldsymbol{W}_R$. Therefore,

$$\begin{aligned}&\boldsymbol{A}_0(\infty) - \boldsymbol{A}_1(\infty) - \boldsymbol{A}_2(\infty) + \boldsymbol{A}_3(\infty)\\ &= 2\left(\boldsymbol{R}_{X_C X_C} + \xi\boldsymbol{R}_W\right)\boldsymbol{\Xi}_{vv}(\infty)\left(\boldsymbol{R}_{X_C X_C} + \xi\boldsymbol{R}_W\right)\\ &\quad -\xi\left(\boldsymbol{R}_{X_C X_C}\boldsymbol{\Xi}_{vv}(\infty)\boldsymbol{R}_W + \boldsymbol{R}_W\boldsymbol{\Xi}_{vv}(\infty)\boldsymbol{R}_{X_C X_C}\right)\\ &\quad + Tr\left(\Delta\boldsymbol{W}\Delta\boldsymbol{W}^T\boldsymbol{R}_{X_C X_C} + \boldsymbol{\Xi}_{vv}(\infty)\boldsymbol{R}_{X_C X_C}\right)\boldsymbol{R}_{X_C X_C}\\ &\quad + \xi^2\tilde{\boldsymbol{R}}_W - \xi^2\boldsymbol{R}_{\boldsymbol{W}}\boldsymbol{\Xi}_{vv}(\infty)\boldsymbol{R}_W.\end{aligned}$$

Assuming μ^2 and ξ^2 are small, we can drop the last term to get

$$\begin{aligned}S(\infty) \approx &\, \mu^2\boldsymbol{D}_x\left[2\tilde{\boldsymbol{R}}_{X_C X_C}\boldsymbol{\Xi}_{vv}(\infty)\tilde{\boldsymbol{R}}_{X_C X_C}\right.\\ &- \xi\left(\boldsymbol{R}_{X_C X_C}\boldsymbol{\Xi}_{vv}(\infty)\boldsymbol{R}_W\right.\\ &\left.+\boldsymbol{R}_W^T\boldsymbol{\Xi}_{vv}(\infty)\boldsymbol{R}_{X_C X_C}\right)\\ &\left.+\left(J_* + \sigma_{\min}^2\right)\boldsymbol{R}_{X_C X_C} + \xi^2\tilde{\boldsymbol{R}}_W\right]\boldsymbol{D}_x^T, \end{aligned}\tag{23}$$

where $J_* = Tr(\boldsymbol{\Xi}_{\boldsymbol{vv}}(\infty)\boldsymbol{R}_{X_C X_C})$ is the steady-state EMSE, and $\sigma_{\min}^2 = Tr(\Delta\boldsymbol{W}\Delta\boldsymbol{W}^T\boldsymbol{R}_{X_C X_C}) + \sigma_\eta^2$ is the minimum energy of the residue. This also gives an upper bound of the MSE since we dropped only the negative terms. Substituting (23) into (18) gives the following equation describing the weight error covariance at the steady state:

$$\begin{aligned}\boldsymbol{\Xi}_{vv}(\infty) \approx &\, \boldsymbol{\Xi}_{vv}(\infty)\\ &-\mu\left(\boldsymbol{D}_x\tilde{\boldsymbol{R}}_{X_C X_C}\boldsymbol{\Xi}_{vv}(\infty) + \boldsymbol{\Xi}_{vv}(\infty)\tilde{\boldsymbol{R}}_{X_C X_C}\boldsymbol{D}_x^T\right)\\ &+2\mu^2\boldsymbol{D}_x\tilde{\boldsymbol{R}}_{X_C X_C}\boldsymbol{\Xi}_{vv}(\infty)\tilde{\boldsymbol{R}}_{X_C X_C}\boldsymbol{D}_x\\ &+\mu^2\boldsymbol{D}_x\left[\left(J_* + \sigma_{\min}^2\right)\boldsymbol{R}_{X_C X_C} + \xi^2\tilde{\boldsymbol{R}}_W\right]\boldsymbol{D}_x^T, \end{aligned}\tag{24}$$

where again we have dropped the negative term to obtain a simplified upper bound of the covariance.

By expressing $\boldsymbol{v}(n)$ in the transformed coordinate: $\boldsymbol{V}(n) = \widetilde{\boldsymbol{U}}^T\boldsymbol{D}_x^{-1/2}\boldsymbol{v}(n)$, (24) can be further simplified to

$$\begin{aligned}\boldsymbol{\Xi}_{\boldsymbol{VV}}(\infty) \approx &\, \boldsymbol{\Xi}_{\boldsymbol{VV}}(\infty) - \mu\widetilde{\Lambda}\boldsymbol{\Xi}_{VV}(\infty) - \mu\boldsymbol{\Xi}_{VV}(\infty)\widetilde{\Lambda}\\ &+\mu^2\Gamma(\infty) + 2\mu^2\widetilde{\Lambda}\boldsymbol{\Xi}_{VV}(\infty)\widetilde{\Lambda}, \end{aligned}\tag{25}$$

where the driving term is $\Gamma(\infty) = (J_* + \sigma_{\min}^2)\Gamma_1 + \Gamma_2$ with $\Gamma_1 = \widetilde{\boldsymbol{U}}\boldsymbol{D}_x^{1/2}\boldsymbol{R}_{X_C X_C}\boldsymbol{D}_x^{1/2}\widetilde{\boldsymbol{U}}^T$ and $\Gamma_2 =$

$\xi^2\widetilde{\boldsymbol{U}}\boldsymbol{D}_x^{1/2}\widetilde{\boldsymbol{R}}_W\boldsymbol{D}_x^{1/2}\widetilde{\boldsymbol{U}}^T$ and we have used the fact $\widetilde{\Lambda}=\widetilde{\boldsymbol{U}}^T\boldsymbol{D}_x^{1/2}\widetilde{\boldsymbol{R}}_{X_CX_C}\boldsymbol{D}_x^{1/2}\widetilde{\boldsymbol{U}}$.

To solve for the EMSE, we first note that the diagonal values of (25) read

$$[\boldsymbol{\Xi}_{\boldsymbol{VV}}(\infty)]_{i,i}\approx[\boldsymbol{\Xi}_{\boldsymbol{VV}}(\infty)]_{i,i}-2\mu\widetilde{\lambda}_i[\boldsymbol{\Xi}_{\boldsymbol{VV}}(\infty)]_{i,i}+\mu^2\cdot\left(2\widetilde{\lambda}_i^2[\boldsymbol{\Xi}_{\boldsymbol{VV}}(\infty)]_{i,i}+\left(J_*+\sigma_{\min}^2\right)[\Gamma_1(\infty)]_{i,i}+[\Gamma_2(\infty)]_{i,i}\right),\tag{26}$$

where $[\Gamma_1(\infty)]_{i,i}=\Gamma_{1,i}$, and $[\Gamma_2(\infty)]_{i,i}=\Gamma_{2,i}$. Then we have the EMSE around the converged solution as:

$$\begin{aligned}J_*&=Tr\Big(\boldsymbol{\Xi}_{\mathbf{vv}}(\infty)\boldsymbol{R}_{X_CX_C}\big)\\&=Tr(\boldsymbol{\Xi}_{\boldsymbol{VV}}(\infty)\widetilde{\boldsymbol{U}}^T\boldsymbol{D}_x^{\frac{1}{2}}\boldsymbol{R}_{X_CX_C}\boldsymbol{D}_x^{\frac{1}{2}}\widetilde{\boldsymbol{U}}\Big).\end{aligned}\tag{27}$$

If the transform can approximately diagonalize $\boldsymbol{R}_{xx}$, then $\overline{\boldsymbol{R}}_{\boldsymbol{xx}}=\boldsymbol{D}_x^{1/2}\widetilde{\boldsymbol{R}}_{X_CX_C}\boldsymbol{D}_x^{1/2}\approx\boldsymbol{I}$, and $\widetilde{\boldsymbol{U}}$ is approximately equal to the identity matrix. Hence,

$$\begin{aligned}\widetilde{\boldsymbol{U}}^T\boldsymbol{D}_x^{\frac{1}{2}}\boldsymbol{R}_{X_CX_C}\boldsymbol{D}_x^{\frac{1}{2}}\widetilde{\boldsymbol{U}}&=\widetilde{\boldsymbol{\Lambda}}-\xi\widetilde{\boldsymbol{U}}^T\boldsymbol{D}_x^{\frac{1}{2}}\boldsymbol{R}_W\boldsymbol{D}_x^{\frac{1}{2}}\widetilde{\boldsymbol{U}}\\&\approx\widetilde{\boldsymbol{\Lambda}}-\xi\boldsymbol{D}_x^{\frac{1}{2}}\boldsymbol{R}_W\boldsymbol{D}_x^{\frac{1}{2}}.\end{aligned}$$

Moreover, for diagonal dominance $\boldsymbol{R}_W$, we further have

$$\begin{aligned}J_*&=Tr\left(\boldsymbol{\Xi}_{\boldsymbol{VV}}(\infty)\left(\widetilde{\Lambda}-\xi\boldsymbol{D}_x^{\frac{1}{2}}\boldsymbol{R}_W\boldsymbol{D}_x^{\frac{1}{2}}\right)\right)\\&\approx\sum_{i=1}^{L}[\boldsymbol{\Xi}_{\boldsymbol{VV}}(\infty)]_{i,i}\left(\widetilde{\lambda}_i-\xi\varepsilon_i^{-1}\boldsymbol{R}_{W\_i,i}\right),\end{aligned}\tag{28}$$

where $\boldsymbol{R}_{W\_i,i}$ is the i-th diagonal value of $\boldsymbol{R}_W$. Solving for $[\boldsymbol{\Xi}_{\boldsymbol{VV}}(\infty)]_{i,i}$ from (26), one gets

$$[\boldsymbol{\Xi}_{\boldsymbol{VV}}(\infty)]_{i,i}=\frac{\mu}{2\widetilde{\lambda}_i(1-\widetilde{\lambda}_i\mu)}\left[\left(J_*+\sigma_{\min}^2\right)\Gamma_{1,i}+\Gamma_{2,i}\right].\tag{29}$$

Consequently, J_* is found to be

$$J_*\cong\frac{\frac{1}{2}\mu\left(\sigma_{\min}^2\phi_{\mathrm{TD}}+\phi_\xi\right)}{\left(1-\frac{1}{2}\mu\phi_{\mathrm{TD}}\right)},\tag{30}$$

where $\phi_{\mathrm{TD}}=\sum_{i=1}^{L}(\Gamma_{1,i}(\widetilde{\lambda}_i-\xi\varepsilon_i^{-1}R_{W\_i,i})/\widetilde{\lambda}_i(1-\widetilde{\lambda}_i\mu))$, $\phi_\xi=\sum_{i=1}^{L}(\Gamma_{2,i}(\widetilde{\lambda}_i-\xi\varepsilon_i^{-1}R_{W\_i,i})/\widetilde{\lambda}_i(1-\widetilde{\lambda}_i\mu))$.

Since the total MSE including the bias is $J_R=E[(\Delta\boldsymbol{W}^T\boldsymbol{X}_C(n))^2]+J_*$, we have from (30) the desired steady-state EMSE as

$$J_R\cong\xi^2\boldsymbol{W}_0^T\boldsymbol{R}_W\boldsymbol{R}_{X_CX_C}^{-1}\boldsymbol{R}_W\boldsymbol{W}_0+\frac{\frac{1}{2}\mu\left(\sigma_{\min}^2\phi_{\mathrm{TD}}+\phi_\xi\right)}{\left(1-\frac{1}{2}\mu\phi_{\mathrm{TD}}\right)}.\tag{31}$$

C. Selection of Regularization Parameter

The regularization parameter plays an important role in the performance of the R-TDNLMS algorithm in terms of steady-state EMSE and convergence rate. Using the performance analysis above, we now derive the regularization parameter ξ to balance the bias and variance components in the MSE deviation from the Wiener solution. For mathematical tractability, we shall assume that the input is white with a variance σ_x^2 and $\boldsymbol{R}_W=\boldsymbol{I}$. Hence $\boldsymbol{D}_x=\varepsilon^{-1}\boldsymbol{I}$ is a diagonal matrix, where ε is the estimate of input signal power σ_x^2. Then, (30) can be simplified to $\phi_{\mathrm{TD}}=(1/\widetilde{\lambda}(1-\widetilde{\lambda}\mu))\Sigma_{i=1}^{L}\Gamma_{1,i}$, where $\widetilde{\lambda}=\widetilde{\lambda}_i=1+\xi\sigma_x^{-2}R_{W\_i,i}\approx1+\xi\varepsilon^{-1}R_{W\_i,i}$ and the second identity follows from the fact that the input power is assumed to be well estimated, i.e. $\sigma_x^2\approx\varepsilon$. Similarly, $\phi_\xi=\Sigma_{i=1}^{L}\Gamma_{2,i}/(\widetilde{\lambda}(1-\widetilde{\lambda}\mu))$. Combining these results, and using $\widetilde{\boldsymbol{U}}\approx\boldsymbol{I}$, (31) can be approximated as

$$\begin{aligned}J_R&\approx\xi^2\sigma_x^{-2}\|\boldsymbol{W}_0\|_2^2+\frac{\frac{1}{2}\mu\left(\sigma_{\min}^2L+\xi^2\sigma_x^{-2}\|\boldsymbol{W}_0\|_2^2\right)}{\widetilde{\lambda}(1-\widetilde{\lambda}\mu)}\\&=\xi^2\sigma_x^{-2}\|\boldsymbol{W}_0\|_2^2\left(1+\frac{\frac{1}{2}\mu}{\widetilde{\lambda}(1-\widetilde{\lambda}\mu)}\right)+\frac{\frac{1}{2}\mu\sigma_{\min}^2L}{\widetilde{\lambda}(1-\widetilde{\lambda}\mu)},\end{aligned}\tag{32}$$

where $\|\boldsymbol{W}_0\|_2^2$ is the norm of the system impulse response which is usually assumed to be known a priori [18]. It can be seen that the first and second terms on the right hand side correspond, respectively, to the bias and variance of MSE. Moreover, $\sigma_{\min}^2=\sigma_\eta^2+\sigma_x^2\|\Delta\boldsymbol{W}\|_2^2$, where $\Delta\boldsymbol{W}\approx\xi\boldsymbol{R}_{X_CX_C}^{-1}\boldsymbol{R}_W\boldsymbol{W}_0$. Therefore,

$$J_R\approx\xi^2\sigma_x^{-2}\|\boldsymbol{W}_0\|_2^2\left(1+\frac{\frac{1}{2}\mu(1+L)}{\widetilde{\lambda}(1-\widetilde{\lambda}\mu)}\right)+\frac{\frac{1}{2}\mu\sigma_\eta^2L}{\widetilde{\lambda}(1-\widetilde{\lambda}\mu)}.\tag{33}$$

It can be seen that the first and second term on the right hand side corresponds, respectively, to the variance and bias of the MSE. In order to obtain a balanced performance in practical applications, we propose to choose ξ so that the two terms are equal to each other. Consequently, the desired regularization parameter satisfies

$$\left(1+\frac{\frac{1}{2}\mu(1+L)}{(1+\xi\varepsilon^{-1})(1-\mu(1+\xi\varepsilon^{-1}))}\right)\xi^2=\frac{\frac{\frac{1}{2}\mu L\sigma_\eta^2\sigma_x^2}{\|\boldsymbol{W}_0\|_2^2}}{(1+\xi\varepsilon^{-1})(1-\mu(1+\xi\varepsilon^{-1}))},$$

which upon solving yields

$$\xi_{opt}\approx\sqrt{\frac{\frac{1}{2}\mu L\sigma_\eta^2\sigma_x^2}{\left[\|\boldsymbol{W}_0\|_2^2\left(1+\frac{1}{2}\mu(1+L)\right)\right]}}\tag{34}$$

Therefore, by estimating the input signal power online with a given prior noise power, the regularization parameter can be adjusted automatically, which yields the proposed variable regularized TDNLMS algorithm.

IV. Simulation Results

Computer simulations are conducted to evaluate the convergence behavior of the proposed algorithm and verify the analytical results obtained in section III. As a comparison, we also consider the conventional TDNLMS algorithm. The DCT transformation is employed due to its wide usage and efficiency in practice. The power of the input element is estimated recursively as $\varepsilon_i(n)=(1-\alpha_\varepsilon)\varepsilon_i(n-1)+\alpha_\varepsilon X_{C,i}^2(n)+\varepsilon$ with $\alpha_\varepsilon=0.01$ and ε either a small constant for stationary input or $\xi(n)[\boldsymbol{R}_W]_{i,i}$, as shown in (8), for power-varying input. All simulations are averaged over 100 independent runs if not specified.

TABLE I
EXPERIMENTAL EMSE RESULTS OF PROPOSED REGULARIZATION PARAMETER FOR FIRST ORDER AR INPUT

SNR (dB)	L	15						50					
	μ	0.007			0.03			0.001			0.005		
0	ξ	$0.1\xi_{opt}$	ξ_{opt}	$5\xi_{opt}$	$0.1\xi_{opt}$	ξ_{opt}	$5\xi_{opt}$	$0.1\xi_{opt}$	ξ_{opt}	$5\xi_{opt}$	$0.1\xi_{opt}$	ξ_{opt}	$5\xi_{opt}$
		0.07	0.68	3.4	0.13	1.3	6.4	0.06	0.6	3	0.12	1.2	6.2
	EMSE	4.2	2.4	8.4	10.2	7.5	17	14.2	11.0	18.9	19.7	15.9	20.6
10	ξ	$0.1\xi_{opt}$	ξ_{opt}	$5\xi_{opt}$	$0.1\xi_{opt}$	ξ_{opt}	$5\xi_{opt}$	$0.1\xi_{opt}$	ξ_{opt}	$5\xi_{opt}$	$0.1\xi_{opt}$	ξ_{opt}	$5\xi_{opt}$
		0.02	0.22	1.1	0.04	0.37	1.8	0.02	0.2	0.99	0.04	0.38	1.9
	EMSE	-4.6	-5.6	0.7	1.2	-1.4	5.6	4.8	2.3	9.3	11.1	8.0	13.7
20	ξ	$0.1\xi_{opt}$	ξ_{opt}	$5\xi_{opt}$	$0.1\xi_{opt}$	ξ_{opt}	$5\xi_{opt}$	$0.1\xi_{opt}$	ξ_{opt}	$5\xi_{opt}$	$0.1\xi_{opt}$	ξ_{opt}	$5\xi_{opt}$
		0.007	0.07	0.34	0.013	0.13	0.64	0.006	0.06	0.3	0.012	0.12	0.62
	EMSE	-14.0	-14.2	-8.3	-7.3	-9.0	-2.1	-4.9	-5.0	0.1	2.0	0.4	5.5

A. Experiment 1: Colored Gaussian Input

This experiment is based on the system identification model in (1). The following first order autoregressive (AR) process is employed as the input: $x(n) = 0.9x(n-1) + g(n)$, where $g(n)$ is a zero-mean and white Gaussian noise. The unknown system to be estimated is an L-order FIR filter. Different signal-to-noise ratios (SNRs) at the system output are used to examine the effect of the regularization on the proposed algorithm under different conditions. First, simulation results of the steady-state EMSE for the R-TDNLMS algorithm using different regularization parameters and step-sizes are compared in Table I. In the simulations, the SNR tested are 0 dB, 10 dB and 20 dB and we evaluate two different filter lengths: $L = 15$ and 50. The step-sizes for $L = 15$ are $\mu = 0.007$ and 0.03 whereas the step-sizes for $L = 50$ are $\mu = 0.001$ and 0.005. The regularization parameters are calculated according to (34) and $\xi = 0.1\xi_{opt}$, ξ_{opt} and $5\xi_{opt}$ are used as comparison. ε is chosen to be 0 since the power of the input element is constant. It can be seen from Table I that the R-TDNLMS algorithm using $\xi = \xi_{opt}$ achieves the lowest steady-state EMSE values compared to those using $\xi = 0.1\xi_{opt}$ and $5\xi_{opt}$ in each case. This suggests that with proper prior information such as the noise power and the norm of the FIR system, (34) is an effective method for selecting the regularization parameter to reduce the MSE.

To further examine the performance of the proposed VR-TDNLMS algorithm, the convergence curves for EMSE are compared to the conventional TDNLMS algorithm in Fig. 1. Moreover, to verify the analysis in Section III, the theoretical predictions of the steady-state EMSE are compared with simulation results. The settings are $L = 15$, $\xi = \xi_{opt}$ with $\mathrm{SNR} = 0$ dB, 10 dB and 20 dB. To see the effect of regularization, we use the same step-size for both algorithms, i.e. $\mu = 0.007$. It can be seen that the VR-TDNLMS algorithm generally converges faster and to a lower steady-state EMSE than the conventional TDNLMS algorithm. The improvement is more significant when the SNR is low as expected. The estimated steady-state EMSE also agrees well with the simulation results. Therefore, the EMSE at (33) given the regularization parameter in (34) gives an expression relating the EMSE with a desired step-size. Hence, it is possible to select a step-size to achieve a given EMSE. We shall further illustrate this approach in the design and implementation of an ANC system in sub-section C below.

B. Experiment 2: Application to Acoustic System Identification

In this experiment, we compare the performance of both the L_1 and L_2 based VR-TDNLMS (L_1-VR-TDNLMS and L_2-VR-TDNLMS) algorithms with the conventional TDNLMS in an acoustic system identification problem. The input is a segment of audio signal with a sampling rate of 8 kHz as shown in Fig. 2(b) and the SNR is chosen as 15 dB. The unknown system is shown in Fig. 2(a). It is used to simulate the impulse response inside an enclosure such as a vehicle and the filter length is $L = 200$.

The step-size for the algorithms is 0.005. For the L_2-VR-TDNLMS algorithm, the regularization parameter is adaptively updated based on (34) as follows:

$$\xi_{L_2}(n) = \overline{\sigma}_x^2 \sqrt{\frac{\frac{1}{2}\mu L\left(\frac{\sigma_\eta^2}{\sigma_x^2(n)}\right)}{\left[\|\boldsymbol{W}_0\|_2^2\left(1+\frac{1}{2}\mu(1+L)\right)\right]}}, \tag{35}$$

where $\overline{\sigma}_x^2$ is an estimate of the averaged input power and $\sigma_x^2(n) = \lambda\sigma_x^2(n-1) + (1-\lambda)x^2(n)$ is the ensemble average of the input power at time instant n, estimated by using a forgetting factor $\lambda = 0.99$. On the other hand, for the L_1-VR-TDNLMS, we have $\xi_{L_1}(n) = \xi_{L2}(n)|\boldsymbol{W}_0|$. Thus, the regularization parameter can be simply derived as [14]

$$\xi_{L_1}(n) = \overline{\sigma}_x^2 \sqrt{\frac{\frac{1}{2}\mu L\left(\frac{\sigma_\eta^2}{\sigma_x^2(n)}\right)}{\left(1+\frac{1}{2}\mu(1+L)\right)}}. \tag{36}$$

Consequently, ε is chosen to be $\xi_l(n)[\boldsymbol{R}_W]_{i,i}$, $l = L_1, L_2$. Since in many practical applications, exact prior information about the system, such as the averaged input and noise power, is not exactly known but an upper bound of these quantities may be estimated, the sensitivity of the above choice of $\xi_l(n)$ is also tested by using $0.1\xi_l(n)$ in the VR-TDNLMS algorithm as a comparison. The learning curves of EMSE of different algorithms are shown in Fig. 2(c). It can be seen that the TDNLMS algorithm is very sensitive to the input signal when the input power is varying considerably. The VR-TDNLMS algorithms, on the other hand, have a high immunity to variation in input power and they achieve much lower EMSE values compared to the conventional TDNLMS algorithm. The performance of L_1 and L_2 regularizations seem to be similar in this case. For L_1 regularization, we also examine the use of multiple updates at each time instant. This is because the use of the previous weight vector in the weighted L_2 regularization is only an approximate solution of the L_1 regularization problem. To obtain a better approximation, we perform one to three updates at each time instant using the first order AR process as input. It can be seen that for one single update, the performance is similar

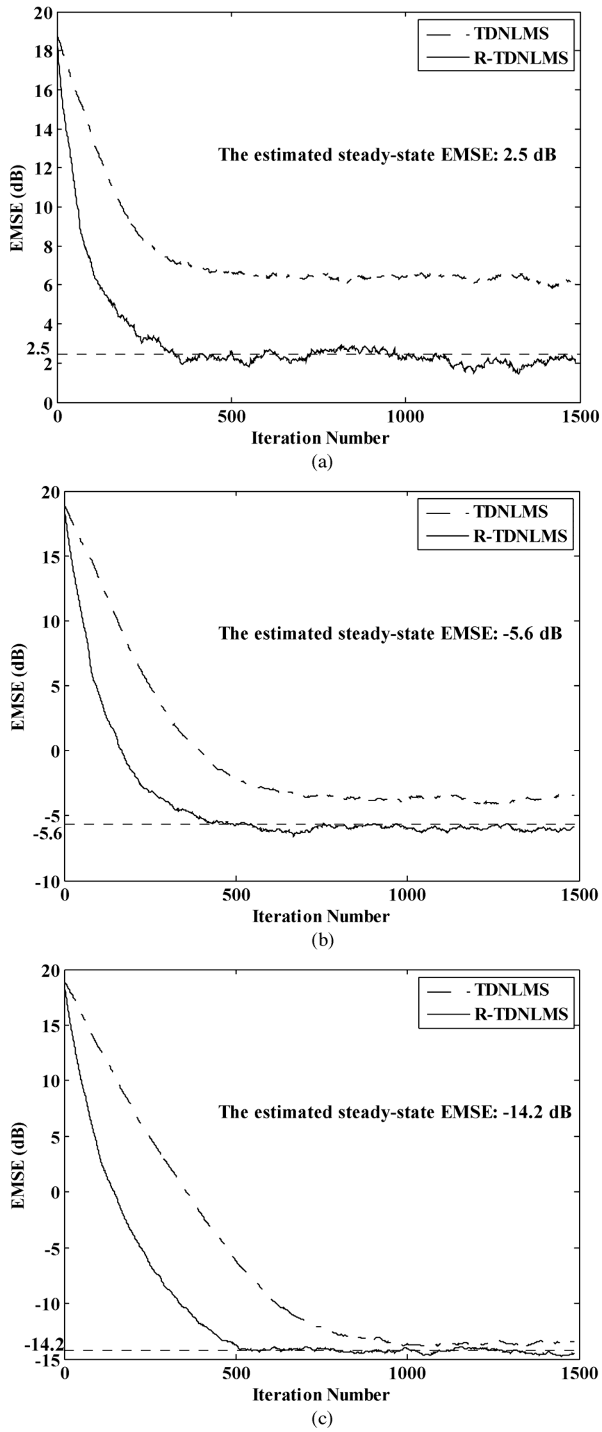


Fig. 1. Learning curves of EMSE for the time-invariant channel identification problem with first-order AR input in experiment 1 at $\mathrm{SNR} = (\mathrm{a})0$ dB (b) 10 dB and (c) 20 dB. $L = 15$.

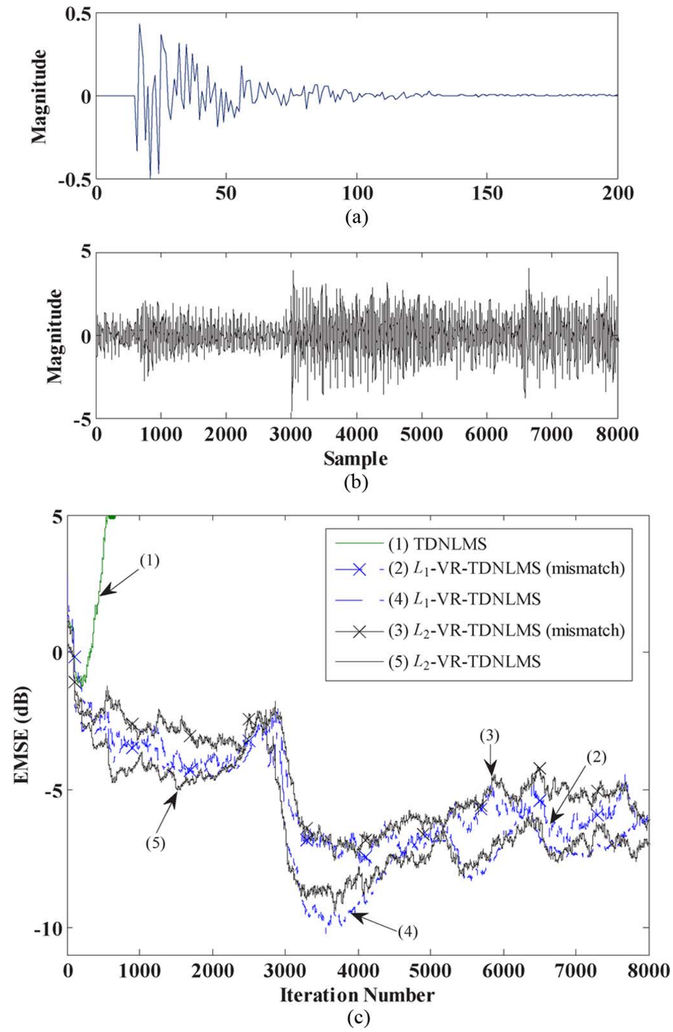


Fig. 2. Learning curves of EMSE (c) for the time-invariant channel (a) identification problem with music input (b) in experiment 2 at $\mathrm{SNR} = 15$ dB. $L = 200$.

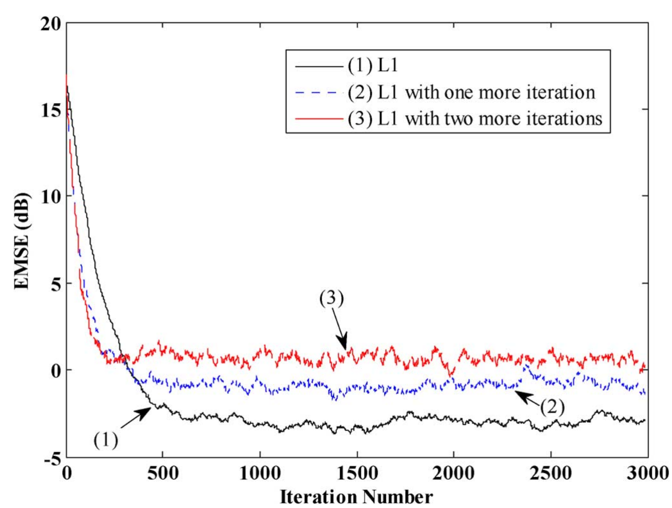


Fig. 3. Learning curves of EMSE for the L_1-VR-TDNLMS with different iteration numbers in experiment 2 at $\mathrm{SNR} = 15$ dB. The input is the first order AR process. $L = 100$.

to the L_2 norm as shown in Fig. 2(c). As the number of iteration increases, improved convergence rate is observed whereas the EMSE is degraded as shown in Fig. 3. This is because the L_1 regularization assumes that the transformed coefficients are sparse and hence the effective number of coefficients to be estimated is smaller. This explains why the initial convergence is faster. On the other hand, since the L_1 is biased as the impulse response is not sparse, hence it exhibits a bias at the steady state, giving rise to higher EMSE. If the impulse response is indeed sparse, better EMSE will be observed. Therefore, the two regularization methods should be properly chosen to suit different

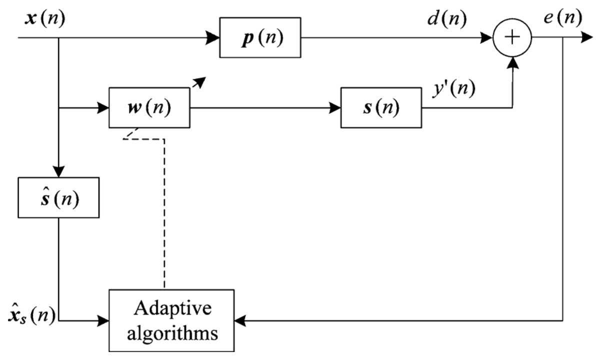


Fig. 4. The diagram of an ANC system.

applications. In the mismatched case, where the regularization may be insufficient, the performance of the VR-TDNLMS algorithms may be slightly degraded.

C. Experiment 3: Application to ANC

In this experiment, an ANC system as shown in Fig. 4 is considered [19]. The impulse responses of the primary and secondary paths are denoted by $\{p_k(n), k = 1,2,\ldots,L_p\}$ where k denotes the length of the path and $\{s_k(n), k = 1,2,\ldots,L_s\}$, respectively. An error microphone is used to pick up the residual signal $e(n)$ to be minimized. Thus after cascading with $\{s_k(n)\}$, the ANC controller $\boldsymbol{w}(n) = [w_1(n),\ldots,w_L(n)]^T$ approximates $\{-p_k(n)\}$ so that the undesirable contribution from the noise source $\{x(n)\}$ is minimized. Since $\{s_k(n)\}$ is unknown, it is replaced by its estimate $\{\hat{s}_k(n)\}$ and the input to the adaptive filter is $\hat{x}_s(n) = x(n)^*\hat{s}_k(n)$, where "*" stands for discrete time convolution. Since $\{x(n)\}$ is filtered by $\{\hat{s}_k(n)\}$, the resultant algorithm is called the Filtered-x (Fx)-based algorithm.

We now consider the application of the proposed VR-TDNLMS algorithm for estimating the Wiener solution of the ANC controller. More precisely, the coefficient vector of the FxTDNLMS algorithm is updated as

$$\boldsymbol{W}(n+1) = \boldsymbol{W}(n) - \mu\boldsymbol{D}_x\boldsymbol{C}\hat{\boldsymbol{x}}_s(n)e(n), \tag{37}$$

and for the proposed variable regularized FxTDNLMS (VR-FxTDNLMS) algorithm, it is given by

$$\boldsymbol{W}(n+1) = (\boldsymbol{I} - \mu\xi(n)\boldsymbol{D}_x\boldsymbol{R}_W)\boldsymbol{W}(n) - \mu\boldsymbol{D}_x\boldsymbol{C}\hat{\boldsymbol{x}}_s(n)e(n), \tag{38}$$

where $\hat{\boldsymbol{x}}_s(n) = [\hat{x}_s(n), \hat{x}_s(n-1),\ldots,\hat{x}_s(n-L_w+1)]^T$.

Since the secondary path is usually not known exactly, the system matrix in the Wiener solution of the Fx-based algorithms is in general asymmetric [20]. Therefore, the ANC system may run into instability [21] because the eigenvalues of the covariance matrix may be negative. This can be improved by better secondary path modeling [20], [22]. In order to solve this problem, regularization techniques can also be employed to improve the robustness of the ANC controller [23], [24] since the additional diagonal term in (12) will help to improve the positive-definiteness of the system matrix. Moreover, the VR-FxTDNLMS algorithm above is a good alternative to the existing regularized FxLMS (R-FxLMS) algorithm [23] because of its fast response and robustness to errors in the secondary-path modeling. In this simulation, the FxTDNLMS, R-FxLMS in [23] and the L_2 based VR-FxTDNLMS algorithms are evaluated.

TABLE II
STABILITY CONDITION

m	$0.1\mu_{\max}$	$\mu_{\max}$	$2\mu_{\max}$	$10\mu_{\max}$	$20\mu_{\max}$
	Short path				
White	0	0	1	100	100
AR	0	0	0	1	100
	Long path				
White	0	0	1	100	100
AR	0	0	1	100	100

m: indicates the divergences for 100 times of Monte Carlo runs. $\mathrm{m} = 0$: no divergence observed; $\mathrm{m} = 1$: diverge at a very small probability (less than 3 times); $\mathrm{m} = 100$: diverge at each trial.

For the VR-FxTDNLMS algorithm, the upper bound of step-size for stability in (22) is firstly examined. In order to examine the accuracy of these upper bounds, different situations have been simulated. Since the results are quite consistent, we only present parts of them to save space. First, a short path length is used. The length of the primary path is 12. The secondary path is estimated offline with a modeling error of -6 dB after normalized by the norm of the true secondary path. Both white and colored Gaussian input sequences are tested. For the colored input, the sequence is generated as a first order AR process as described in Experiment 1. The background noise variance is 0.064. The step-size upper bound is then calculated to be $\mu_{\max} = 0.05$ and $\mu_{\max} = 0.01$, respectively, for the white and colored inputs. Then, we employ different step-sizes, i.e. $\mu = 0.1\mu_{\max}, \mu_{\max}, 2\mu_{\max}, 10\mu_{\max}$ and $20\mu_{\max}$, and examine the stability conditions, which is summarized in Table II. Note, 100 random runs are used to examine if there is any possibility of instability. It is then repeated by using longer path lengths, i.e. the length of the primary path is 100. For the white and colored inputs, $\mu_{\max}$ is calculated to be 0.004 and 0.001, respectively. From the convergence conditions, we expect that the system becomes unstable when the step-size exceeds the upper bounds. It can be seen that the bound is quite accurate, since beyond which the system usually becomes unstable for all or parts of the 100 random runs. It can be seen that the algorithm may still be stable or the probability of divergence is low (less than 3 divergences were observed in 100 runs) when the step-size is close to the bound, say 2 times of $\mu_{\max}$. Overall, it can be seen that the theoretical prediction provides valuable information about the selection of step-sizes for stable operation.

In the second experiment, we shall propose a design procedure for ANC systems based on the theoretical analyses obtained in Section III and illustrate its usefulness by a design example.

The settings are identical to the above example for the longer path length. The desired steady-state EMSE J_* for the ANC controller is set to be 0.06 (around -12 dB) and 0.015 (around -18 dB), respectively, for white and colored inputs. Given the desired EMSE, we have to determine two unknown parameters μ and ξ. One solution is to solve it iteratively from some initial guess of the step-size μ. Substituting the initial value of μ into (34), one gets an updated optimal regularization parameter ξ. We can then update μ from the updated ξ according to (30) and

TABLE III
PROPOSED DESIGN PROCEDURE FOR ANC SYSTEMS

Given the desired steady-state EMSE for the ANC controller, ξ_w, as well as the prior knowledge or estimate of the input power and noise variances, σ_η^2.

Step 1 Calculate the regularization parameter from (34):

$$\xi \approx \sqrt{\tfrac{1}{2}\hat{\mu}L\sigma_\eta^2\sigma_x^2 / (\| \boldsymbol{W}_0 \|_2^2 \, (1+\tfrac{1}{2}\hat{\mu}(1+L)))},$$

where $\hat{\mu} = 2J_* / (\sigma_e^2 \phi_{TD})$ is a tentative step-size and $\sigma_e^2 = \sigma_\eta^2 + \sigma_w^2$. σ_w^2 is the variance of the system residue $x(n)*p_k(n)+\boldsymbol{w}*x_s(n)$ obtained numerically.

Step 2 Calculate the step-size of the ANC controller μ from (30) using the small step size (SSS) approximation:

$$\mu \cong \frac{2J_*}{(\sigma_{\min}^2 \phi_{TD} + \phi_\xi)},$$

where $\sigma_{\min}^2 \approx \sigma_\eta^2$ (23), $\phi_{TD} \approx \sum_{i=1}^{L} (1-\xi\varepsilon_i^{-1})/(1+\xi\varepsilon_i^{-1})$, and $\phi_\xi \approx \sum_{i=1}^{L} \xi^2 \varepsilon_i^{-1} W_{R\_i}^2 (1-\xi\varepsilon_i^{-1})/(1+\xi\varepsilon_i^{-1})$ in (30), with W_{R_i} being the i-th element of $\boldsymbol{W}_R$.

Step 3 Check the stability conditions for the ANC controller and the secondary-path estimator in (22). If this condition is violated, a smaller μ should be chosen.

---

so on. Our simulation results show that, given an appropriate initial guess, one time of iteration could give a satisfactory result. This procedure is summarized in Table III for the ANC design.

Usually, a small step-size is used to achieve a good performance. Thus, an appropriate initial guess of μ can be calculated from $\hat{\mu} \approx 2J_*/(\sigma_e^2 \phi_{\mathrm{TD}})$, where $\sigma_e^2 = \sigma_\eta^2 + \sigma_w^2$ and σ_w^2 is the variance of the system residue $x(n)^* p_k(n) + \boldsymbol{w}^* x_s(n)$ [20], which can be obtained numerically. Using this initial step-size, an approximate regularization parameter can be calculated (Step 1 in Table III). Under the small step-size assumption, the step-size for the ANC controller can be calculated from (30) to be $\mu \approx 2J_*/(\sigma_{\min}^2 \phi_{\mathrm{TD}} + \phi_\xi)$, where the term $(1/2)\mu\phi_{\mathrm{TD}}$ in the denominator is ignored since μ is small. In addition, the following approximation for ϕ_{TD} and ϕ_ξ can be used, i.e. $\phi_{\mathrm{TD}} \approx \sum_{i=1}^{L}(1-\xi\varepsilon_i^{-1})/(1+\xi\varepsilon_i^{-1})$ and $\phi_\xi \approx \sum_{i=1}^{L} \xi^2 \varepsilon_i^{-1} W_{R\_i}^2 (1-\xi\varepsilon_i^{-1})/(1+\xi\varepsilon_i^{-1})$, with W_{R_i} being the i-th element of $\boldsymbol{W}_R$. Consequently, the desired step-sizes are $\mu =$0.0025 and 0.0008, respectively, for white and colored inputs. Finally, applying the two step-sizes to (22), we found that the stability constraints are satisfied (Step 3).

The calculated step-sizes are used in the ANC system for the VR-FxTDNLMS algorithm. The regularization parameter is chosen according to (35). The EMSE curves, according to (5), are plotted in Fig. 5. It shows that, the VR-FxTDNLMS algorithm using the two designed step-sizes achieves approximately steady-state EMSEs at -12 dB and -18 dB, respectively for the white and colored inputs, which are close to the desired values. This illustrates the effectiveness of the proposed design procedure for ANC systems.

Moreover, the mismatched case with a smaller regularization parameter of $0.1\xi(n)$, is also tested as a comparison. Their convergence performance is also compared with the R-FxLMS algorithm under the same setting. For R-FxLMS, the step-sizes and regularization parameters are tuned to arrive at a similar steady-state EMSE. The resultant step-sizes for white and colored inputs are chosen to be 0.002 and 0.0005, respectively, and

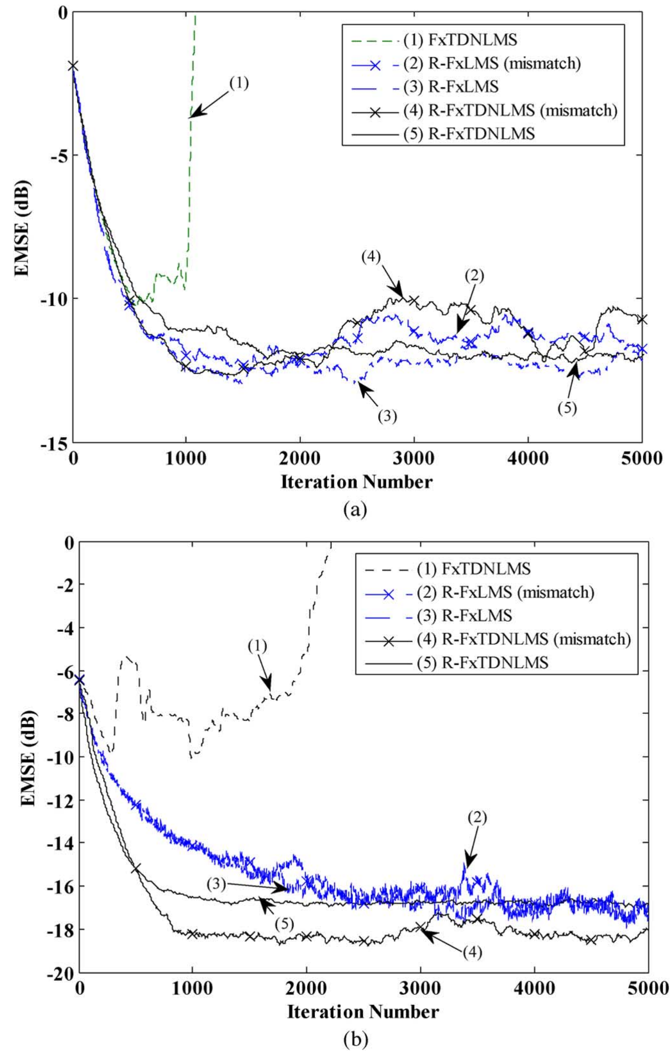


Fig. 5. Performance comparison of various algorithms for the ANC problem with (a) white input and (b) colored input. The background noise variance is 0.064. The modeling error is -6 dB.

the regularization parameter is found to be 0.1. As a comparison, a mismatched regularization parameter of $\xi = 0.01$ is also used in the R-FxLMS algorithm.

Fig. 5 shows the EMSE learning curves of the various algorithms. It can be seen that for both white and colored inputs, the FxTDNLMS algorithm diverges since the modeling error of the secondary-path results in a singular system matrix. The regularized algorithms, on the other hand, show improved robustness under such circumstances. Moreover, the transform domain algorithm accelerates the convergence speed for the colored input. For the colored input, the FxTDNLMS algorithm with $0.1\xi(n)$ seems to achieve a lower steady-state EMSE. This is probably because the modeling error of the secondary path is not taken into account when selecting the regularization parameter, though the latter contributes effectively to improve the positive definiteness of the resultant system.

V. CONCLUSIONS

A new R-TDNLMS algorithm and its mean and mean square convergence performances have been presented. The proposed algorithm extends the conventional TDNLMS algorithm by imposing a regularization term on the filter coefficients to reduce the variance of estimators due to the lacking of excitation or in the presence of modeling errors. Difference equations describing the mean and mean square convergence behaviors of this algorithm are derived so as to characterize its convergence condition and steady-state EMSE. Based on this analysis, a new formula to select the regularization parameter for white Gaussian inputs is proposed, which leads to a new VR-TDNLMS algorithm. The improved convergence performance, steady-state EMSE and robustness to power-varying inputs of the VR-TDNLMS algorithm and the effectiveness of the theoretical analysis are verified by computer simulations. The application of the proposed VR-TDNLMS algorithm to the design and implementation of acoustic system identification and ANC systems are also illustrated by design examples. Improvements over traditional TDNLMS algorithms at low excitation or in the presence of modeling errors are observed. Moreover, the theoretical analysis provides simple design formulas for achieving a given EMSE and step-size bound for stable operation.

REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, 4th ed. Englewood Cliffs, NJ: Prentice-Hall, 2001.

[2] B. Widrow, J. M. McCool, M. G. Larimore, and C. R. Johnson, Jr., "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. IEEE*, vol. 64, no. 8, pp. 1151–1162, Aug. 1976.

[3] J. I. Nagumo and A. Noda, "A learning method for system identification," *IEEE Trans. Autom. Contr.*, vol. AC-12, no. 3, pp. 282–287, Jun. 1967.

[4] S. Narayan, A. M. Peterson, and M. J. Narasimha, "Transform domain LMS algorithm," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. ASSP-31, no. 3, pp. 609–615, Jun. 1983.

[5] J. C. Lee and C. K. Un, "Performance of transform-domain LMS adaptive digital filters," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 3, pp. 499–510, Jun. 1986.

[6] S. C. Chan and Y. Zhou, "On the convergence analysis of the transform domain normalized LMS and related M-estimate algorithms," in *Proc. IEEE APCCAS '08*, Macao, Nov. 30–Dec. 3 2008.

[7] F. Beaufays, "Transform-domain adaptive filters: An analytical approach," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 422–431, Feb. 1995.

[8] M. P. Nowak and B. D. Van Veen, "A constrained transform domain adaptive IIR filter structure for active noise control," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 422–431, Feb. 1995.

[9] S. C. Chan, Y. J. Chu, and Z. G. Zhang, "A new regularized transform-domain NLMS adaptive filtering algorithm," in *Proc. APCCAS '10*, Kuala Lumpur, Malaysia, Dec. 6–9, 2010, pp. 696–699.

[10] S. L. Gay, "Dynamically regularized fast RLS with application to echo cancellation," in *Proc. ICASSP '96*, vol. 2, pp. 957–960.

[11] S. C. Chan, Y. J. Chu, and Z. G. Zhang, "A new regularized QRD recursive least M-estimate algorithm: Performance analysis and applications," in *Proc. ICGCS '10*, Shanghai, China, Jun. 21–23, 2010.

[12] Y. Chen, Y. Gu, and A. O. Hero, "Sparse LMS for system identification," in *Proc. ICASSP '09*, Taipei, Taiwan, Apr. 19–24, 2009, pp. 3125–3128.

[13] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Amer. Stat. Assoc.*, vol. 96, no. 456, pp. 1348–1360, Dec. 2001.

[14] S. C. Chan, Z. G. Zhang, and Y. J. Chu, "A new transform-domain regularized recursive least M-estimate algorithm for robust linear estimation," *IEEE Trans. Circuits Syst. II*, vol. 58, no. 2, pp. 120–128, Feb. 2011.

[15] Z. G. Zhang, S. C. Chan, Y. Zhou, and Y. Hu, "Robust linear estimation using M-estimation and weighted L1 regularization: Model selection and Recursive Implementation," in *Proc. IEEE ISCAS '09*, Taipei, Taiwan, May 24–27, 2009, pp. 1193–1196.

[16] S. Boyd, L. Vandenberghe, and M. Grant, "Efficient convex optimization for engineering design," in *Proc. IFAC Symp. Robust Control Design.*, Rio de Janeiro, Brazil, Sep. 1994.

[17] A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-33, no. 1, pp. 222–230, Feb. 1985.

[18] T. V. Waterschoot, G. Rombouts, and M. Moonen, "Optimally regularized adaptive filtering algorithms for room acoustic signal enhancement," *Signal Process.*, vol. 88, pp. 594–611, 2008.

[19] S. M. Kuo and D. R. Morgan, *Active Noise Control Systems—Algorithms and DSP Implementations*. New York: Wiley, 1996.

[20] S. C. Chan and Y. J. Chu, "Performance analysis and design of FxLMS algorithm in broadband ANC system with online secondary-path modeling," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 3, pp. 982–993, Mar. 2012.

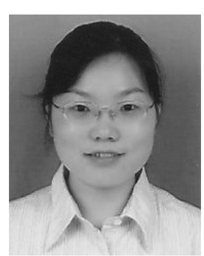
[21] R. Fraanje, M. Verhaegen, and S. J. Elliott, "Robustness of the filtered-x LMS algorithm-Part I: Necessary conditions for convergence and the asymptotic pseudospectrum of Toeplitz matrices," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4029–4037, Aug. 2007.

[22] S. Boroujeny and M. Eshghi, "Online secondary path modeling in active noise control system using PBS-LMS algorithm," in *Proc. IEEE ICSP '10*, Bejing, China, Oct. 24–28, 2010, pp. 2600–2603.

[23] O. J. Tobias and R. Seara, "Leaky-FXLMS algorithm: Stochastic analysis for Gaussian data and secondary path modeling error," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 6, pp. 1217–1230, Nov. 2005.

[24] R. Fraanje, S. J. Elliott, and M. Verhaegen, "Robustness of the filtered-x LMS algorithm-Part II: Robustness enhancement by minimal regularization for norm bounded uncertainty," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4038–4047, Aug. 2007.

S. C. Chan (S'87–M'92) received the B.Sc. (Eng) and Ph.D. degrees from The University of Hong Kong, Pokfulam, in 1986 and 1992, respectively. Since 1994, he has been with the Department of Electrical and Electronic Engineering, the University of Hong Kong, where he is currently a professor. His research interests include fast transform algorithms, filter design and realization, multirate and biomedical signal processing, communications and array signal processing, high-speed A/D converter architecture, bioinformatics, smart grid image-based rendering. He is currently a member of the Digital Signal Processing Technical Committee of the IEEE Circuits and Systems Society, and associate editors of *Journal of Signal Processing Systems* and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II. He was the chair of the IEEE Hong Kong Chapter of Signal Processing in 2000–2002, an organizing committee member of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, the 2010 International Conference on Image Processing, and an associate editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I from 2008 to 2009.



Y. J. Chu received the B. S. and M. S. degree in acoustics from the Nanjing University, Nanjing, China, in 2005 and 2008, respectively. She is currently working towards the Ph.D. with the electrical and electronics engineering department of The University of Hong Kong, Hong Kong. In 2006, she served as a research assistant in the department of building service engineering of the Polytechnic University of Hong Kong, Hong Kong. Her current research interest is the analysis, design and applications of robust adaptive filtering algorithms and recursive detection methods.

Z. G. Zhang (S'05–M'07) received the B.Sc. degree in electrical and electronic engineering from Tianjin University, Tianjin, China, in 2000, the M.Eng. degree in electrical and electronic engineering from the University of Science and Technology of China, Hefei, China, in 2003, and the Ph.D. degree from the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, in 2008. He is currently a Research Assistant Professor with The University of Hong Kong. His research interests include biomedical signal processing, neural engineering and computational neuroscience.